7057-106

U.S. ARMY INTELLIGENCE CENTER AND SCHOOL SOFTWARE ANALYSIS AND MANAGEMENT SYSTEM

OPTIMAL FIX ESTIMATORS AND A VARIATION ON THE CRAMER RAO LOWER BOUND

Institute for Descision Sciences

03 September 1987

National Aeronautics and
Space Administration

JPL

JET PROPULSION LABORATORY
California Institute of Technology
Pasadena, California

JPL D-4718
ALGO_PUB_0124

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: ALGO PUB 0124

2. GOVT ACCESSION NO.:

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE *(and Subtitle)*: Optimal Fix Estimators and a Variation on the Cramer Rao Lower Bound

5. TYPE OF REPORT & PERIOD COVERED: FINAL

6. PERFORMING ORG. REPORT NUMBER: D-4718

7. AUTHOR(s): Institute for Decision Sciences, Claremont McKenna College

8. CONTRACT OR GRANT NUMBER(s): NAS7-918

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Jet Propulsion Laboratory, ATTN: 171-209
California Institute of Technology
4800 Oak Grove, Pasadena, CA 91109

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: RE 182 AMEND #187

11. CONTROLLING OFFICE NAME AND ADDRESS:
Commander, USAICS
ATTN: ATSI-CD-SF
Ft. Huachuca, AZ 85613-7000

12. REPORT DATE: 03 Sep 87

13. NUMBER OF PAGES: 22

14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*:
Commander, USAICS
ATTN: ATSI-CD-SF
Ft. Huachuca, AZ 85613-7000

15. SECURITY CLASS. *(of this report)*: UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: NONE

16. DISTRIBUTION STATEMENT *(of this Report)*:
Approved for Public Dissemination

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*:

18. SUPPLEMENTARY NOTES:
Prepared by Jet Propulsion Laboratory for the US Army Intelligence Center and School's Combat Developer's Support Facility.

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*:
Optimal Location Estimate, Weighted Mean Square, Generalized Cramer Rao Lower Bound, s^2-biased Estimators, Linearization, Optimal, Uniformly Minimum Variance Unbiased Estimator (UMVUE)

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*:
Via an appendix, the Cramer Rao lower bound is generalized: Of all linear estimators of fix (ie. two dimensional) error with s^2 bias, the Least Squares estimator has the smallest error ellipse. The report discusses the assumptions of error independence and the linearization of error model and corresponding small errors. By combining this report with "Two Dimensional Uncorrelated Bias in Fix Algorithms", DTIC #AD- A189473, the "best" fix algorithm is Minimization of Squared Angular Error.

DD FORM 1473, 1 JAN 73 — EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

7057-106

U.S. ARMY INTELLIGENCE CENTER AND SCHOOL
Software Analysis and Management System

Optimal Fix Estimators and a Variation on the Cramer Rao Lower Bound

03 September 1987

Author:

Institute for Decision Sciences

Approval:

James W. Gillis, Subgroup Leader
Algorithm Analysis Subgroup

Edward J. Records, Supervisor
USAMS Task

Concur:

A. F. Ellman, Manager
Ground Data Systems Section

Fred Vote, Manager
Advanced Tactical Systems

JET PROPULSION LABORATORY
California Institute of Technology
Pasadena, California

JPL D-4718

PREFACE

The work described in this publication was performed by the Institute For Decision Sciences (IDS) under contract to the Jet Propulsion Laboratory, an operating division of the California Institute of Technology. This activity is sponsored by the Jet Propulsion Laboratory under contract NAS7-918, RE182, A187 with the National Aeronautics and Space Administration, for the United States Army Intelligence Center and School.

This specific work was performed in accordance with the FY-87 statement of work (SOW #2).

CONTENTS

SUMMARY

The fix estimators MARC (Mathematical Analysis Research Corportation) has examined in fielded systems have a property which implies that these fix estimators approach optimality. Explaining the meaning of this statement and the qualifications that go with it is the purpose of the main body of this report. Proofs needed are included in a Math Appendix.

A list of topics covered in the individual sections of this report follows:

I. OPTIMAL ESTIMATORS IN STATISTICS -

- A. Optimality has no unique definition
- B. Bias
- C. Variance
- D. Uniformly Minimum Variance Unbiased Estimators (UMVUEs) as optimal.
- E. Cramer Rao Lower Bound as a means of finding UMVUEs.
- F. Differences between the fixing case and the standard case: In the fixing case,
  1. no unbiased estimators exist
  2. bearing measurements are not identically distributed
  3. estimating location involves determining more than one parameter.

II. FIX MODELING ASSUMPTIONS -

- A. Independent Normally distributed bearing error with common standard deviation, σ.
- B. Minimization of Squared Angular Error
- C. Small errors and linearization
- D. Discussion of weighting as means of linearizing optimally
- E. Ellipse dependence on linearization
- F. Optimality being relative to the linearized model

III. PROBLEMS WITH FIX MODELING ASSUMPTIONS -

- A. Bias resulting from the non-linear terms
- B. Bearing selection interfering with the fix modeling assumptions
- C. Possibility of dependent bearings
- D. If σ is known, then better models are possible

IV. OPTIMALITY TO WITHIN TERMS OF ORDER σ^2.

A. Bias is of order σ^2

B. Cramer Rao Lower Bound result can be generalized to order σ^2

C. Which methods are order σ^2 optimal?

D. Choosing the best order σ^2 optimal method without considering data storage and computation time.

E. How much does choosing the best order σ^2 optimal method matter?

I. OPTIMAL ESTIMATORS IN STATISTICS

A. Optimality has no unique definition

The estimator which is 'best' in any particular situation depends on the nature of:

1) The information available to be used by the estimator.
2) The overall reward/penalty function which applies, considering the variety of behaviors the estimator may exhibit and the applications to be made of it.

Analysis can only approximate these two considerations. What statistics does do is identify measures of desirable behavior and the means of optimizing these measures given idealized input.

B. Bias

Estimators produce different results depending on the nature of the errors in the observations. The size of the error in the fix depends upon chance. In practice, only one fix is made. It is useful however to imagine averaging all of the fixes that chance might have produced. If these potential fixes are weighted by the probability of their occurrence, then one has the 'expected fix'. If the 'expected fix' turns out to be located at the true emitter, then the estimator would be referred to as unbiased. When the 'expected fix' turns out to be located away from the true emitter, exactly how far apart these two points are becomes an issue of concern. In this case, the estimator is referred to as biased and the distance between the points is called the bias.

Estimators used in practice are biased. The 'expected fix' is short in range of the true for the most commonly used estimator. However, the bias (distance short) is usually small (small in comparison with the variance).

C. Variance

While bias is the distance between the 'expected fix' and the true emitter, variance is a means of measuring the average distance between the 'expected fix' and individual estimates. If the variance of an estimator is large, any confidence in the accuracy of the estimate is limited. A small variance indicates an estimator which will consistently give similar estimates. However, variance does not include any measure of the distance between the 'expected fix' and the true. It is quite possible to have an estimator with small variance which estimates the location to

be a significant distance from the true position of the emitter owing to bias.

D. Uniformly Minimum Variance Unbiased Estimators as optimal

As was mentioned above, there exists no unique definition of what constitutes an optimal estimator. In part, when defining an optimal estimator, one must consider the constraints of available information. Estimates of both bias and variance fall within these constraints, but minimizing only bias without considering variance (or vice-versa) allows no control over the size of the variance. The optimal method, then, with regards to both bias and variance, is to minimize both. Standard practice has been to minimize the variance of an unbiased estimator, if such an estimator can be found. If such an estimator exists, then it is known as the uniformly minimum variance unbiased estimator (UMVUE).

E. Cramer Rao Lower Bound as a means of finding UMVUEs

The Cramer Rao Inequality provides a lower bound for the variance of an unbiased estimator. Thus, in attempting to determine which is the 'best' unbiased estimator among the set of unbiased estimators, the Cramer Rao Lower Bound can be utilized as a measure. Specifically, if the variance of an unbiased estimator satisfies the Cramer Rao lower bound, then that estimator can be said to have the minimum variance among unbiased estimators (in other words, a UMVUE). Such an estimator would then be optimal with regards to bias and variance.

F. Differences between the fixing case and the standard case

In our case,

1) There are no existing unbiased fix estimators. For the two LOB case, it can be shown that when σ^2 is unknown, it is not possible for an estimator to be unbiased. For three or more LOBs, however, it has not yet been resolved if an unbiased estimator can exist. Even so, the authors of this report are unaware of any existing unbiased estimators.
2) Bearing measurements are not identically distributed due to the fact that they all have different means (expected values).
3) Usually, finding UMVUEs requires estimating only one parameter. Yet, estimating location requires determining both an 'x' and a 'y' parameter in order to estimate location.

II. FIX MODELING ASSUMPTIONS

Optimality is measured against models and hence one must be careful about modeling assumptions when discussing optimality.

A. Bearing error distribution

The assumptions are:

1) Independence - This means that the size and direction of a particular bearing error does not influence the size and direction of another bearing error. An example of where this assumption loses validity is when error is induced by inaccurate determination of True North.

2) Normally distributed error with mean zero - Normality is not absolutely required. Symmetric error with mean zero is the most important property of the normal distribution that is needed.

3) Common standard deviation - Knowledge about differences in the standard deviations of different bearings should be used if available. Unknown differences can be tolerated so long as the difference is not too severe. Different signal to noise ratios at different sensors is an example of what might signify a difference in angular error standard deviation.

B. Minimization of Squared Angular Error

This model assumes that the error is in the bearing angle as opposed to the sensor location. Furthermore, in order to find the best fix, a score is computed by taking the angular error and *squaring* it. Thus, a bearing error three times as large would be nine times as bad. This is one of the consequences of assuming normality. One 'wild' bearing (from another emitter perhaps) can have a very large impact because of this assumption. If a higher power were used, then outliers would determine the fix estimate completely.

C. Small errors and linearization

Fix algorithms take advantage of the fact that angular error is small. Linearization of angular error into spatial error is relatively accurate because angular error is small. If this were not true, then:

1) Fixes would be very biased.

2) The 'Weighted Perpendicular' method (the name used in MARC reports for the most frequently used method of approximating minimization of angular error) would not be a good approximation to the Minimization of Squared Angular Error Method which it is supposed to behave like.

D. <u>Weighting to attain optimal linearization</u>

The 'Weighted Perpendicular' method iteratively reweights data on the basis of the latest estimate of the fix location. The limitations on reweightings of updated fixes is one of the major differences between algorithms (limitations are caused by memory and speed limitations of the computer). The objective of the reweighting is to obtain a solution in terms of the best linearization of the angular error that can be found. Not all algorithms are optimal in this reweighting sense because of memory and speed limitations.

E. <u>Ellipse dependence on linearization</u>

The ellipse comes from a quadratic form. That quadratic form is based on the covariance of a linearization (in the sense of a Taylor Series to the first derivative). If angular errors were not small, then the correct family of curves to use to indicate error regions would be much more difficult to manipulate.

F. <u>The optimality of this report is relative to the linear approximation</u>

The optimality of concern in this report is qualified. It will only assert optimality to within a small second order term. This qualification should be expected, however, as most of fix estimation theory is based on the same qualification.

III. PROBLEMS WITH THE FIX MODELING ASSUMPTIONS

A. Bias resulting from the non-linear terms

After reading the previous section, it should not be surprising that non-linear terms are one of the major problems with fix modeling assumptions. Bias is the most significant impact of non-linear terms. It is the difference between where you would expect a fix to be located on average and where it really is. When non-linear terms are small, bias is not noticeable. This is frequently the case. Nonetheless, bias does exist for all methods except under very special circumstances. The direction and size of bias can vary from method to method. For the 'Weighted Perpendicular' method with three or more bearings, bias can be shown to point short along the range. For other methods, bias can be long.

One of the disadvantages of the 'Weighted Perpendicular' method is that bias does not necessarily get smaller as the sample size increases. In fact, bias may even increase with sample size. If a large amount of data is going to be used in a fix, there are better methods. Unfortunately, the other methods require storing all of this data.

B. Bearing selection model impact

Bearing selection is not part of the fix model as used by the Army systems reviewed by MARC. Other models, such as those which add a uniform background to a truncated normal, may reflect bearing selection issues but MARC has not investigated this aspect of the question. It is clear, however, that bearing selection can interfere with the assumption of normality. There are many possible repercussions of this loss. For example, minimizing the squared error may not be optimal.

C. Dependent bearings

If bearing errors are dependent on one another (for reasons such as a shared error in determination of True North), then fix errors should be expected to be larger than normally predicted. If the dependence is strong enough, then an optimal model would attempt to account for this dependence.

D. Is σ (Angular Error Standard Deviation) known or unknown

The current models are ambiguous with respect to knowing σ^2. If σ^2 is known, then it would be possible to correct for fix algorithm induced bias. Even a known lower bound would yield a lower bound correction. The correction would be suspect if σ^2 is not constant across bearings, however. The methods of σ^2 determination that have been reviewed by the authors of this report raise more questions about underlying models than they resolve.

IV. OPTIMALITY TO WITHIN TERMS OF ORDER σ^2

A. Bias is of order σ^2

Showing that bias must be of at least order σ^2 (when σ^2 is unknown) for particular algorithms is simple. Showing it for all algorithms in general is less so. The two bearing case, however, can be shown to be of order σ^2 for all algorithms. Correcting for the bias of the intersection (the two bearing case) depends on knowing σ^2 and σ^2 cannot even be estimated using two bearings. Since σ^2 can be estimated with three or more bearings, it seems possible that .here might be an order σ^4 algorithm that could be constructed in these cases. The authors of this report know of no such algorithms in use. Problems in the estimation of

σ^2 suggest to the authors of this report that such algorithms would have limited application in any case.

Since bias of order σ^2 exists in most, if not all, fielded systems, it seems reasonable to ask whether fix algorithms are optimal to within a term of order σ^2. That is the purpose of this report. Finding the method with the smallest order σ^2 term is a reasonable follow-up question. This question has already been addressed in part for a broad class of fix algorithms in MARC's Two Dimensional Uncorrelated Bias in Fix Algorithms (12 March 87).

B. Cramer Rao Lower Bound can be generalized

The Cramer Rao Lower Bound is a theorem which gives a bound on how accurate an unbiased estimator can be (where accuracy is measured by the size of the variance). Since there are no unbiased estimators, a generalization of the Cramer Rao Lower Bound is needed. Such a generalization is derived in the Math Appendix. The differences between our version of the Cramer Rao Lower Bound and versions found in Wilk's *Mathematical Statistics* are :

1) The result in this report is for order σ^2 biased estimators.
2) The result in this report is for 2-dimensional estimators instead of one-dimensional estimators. (A partial result for n-dimensional estimators is also in this report.)

C. Which methods are order σ^2 optimal?

All fielded systems that MARC has reviewed are order σ^2 optimal. The only theoretical method which MARC has examined which is not optimal to within order σ^2 is the unweighted version of the Perpendicular Method. The variance of all other methods are essentially the same.

D. The 'best' order σ^2 optimal method

There is a difference in the order σ^2 term of various methods. The impact of this order σ^2 term on bias is much more important than its impact on the variance. Minimization of Squared Angular Error and methods similar to it have a smaller order σ^2 bias than the more commonly used Weighted Perpendicular Method.

Both the Minimization of Squared Angular Error and Weighted Perpendicular Methods can only be used in their intended form as long as bearings are saved. After eliminating bearings, both methods must give way to the suboptimal version of the Weighted Perpendicular. Therefore, the only justifications for use of the Weighted Perpendicular are greater simplicity in the coding of the algorithm and simplicity of computation. The Weighted Perpendicular method also generalizes easily to a 3-dimensional method.

E. How significant is the algorithm induced bias?

It depends on:

1) Sensor accuracy (With very accurate bearings, the bias is insignificant).

2) The number of bearings kept before using the suboptimal approach. With two bearings, there is no difference. With four or more bearings, the difference is of more interest.

3) The range of compass headings over which bearings are taken (excluding bearings taken from much further away than the nearest sensors to the emitter). If the range is too small, then bias can be very significant.

Furthermore, if error originates from other sources than bearing

accuracy such as sensor location error, then both Weighted Perpendicular and Minimization of Angular Error are less than optimal. It is this last consideration that has led the authors of this report to conclude that changing from Weighted Perpendicular to Minimization of Squared Angular Error need only be of concern where location estimates can be shown to have been short in range on average (This is the form of bias which the Weighted Perpendicular exhibits). Even in this case, there may be other causes besides the difference between these two algorithms.

MATH APPENDIX

DEFINITIONS:

Let θ_i = ith piece of observed data (bearing from sensor i in our application).

Let (α1,...,αn) be the true location in n-space (of the emitter in our application).

Let (X1,X2,...,Xn) = (X1($\theta_1,\theta_2,\ldots,\theta_m$),...,Xn($\theta_1,\theta_2,\ldots,\theta_m$)).

Let (X1,...,Xn) be estimators of (α1,...,αn) such that

$$E\{(X1,\ldots,Xn)\} = (\alpha 1,\ldots,\alpha n) + (O(\sigma^2),\ldots,O(\sigma^2))$$

where O(h(σ)) is such that, k = O(h(σ)) means that as σ approaches some limit, k(σ) is dominated by some positive constant multiple of h(σ). (See Olmsted, John M., Real Variables , Apple Century Crofts, Inc., 1956, pg. 169)

Let f($\theta_1,\ldots,\theta_m$;(α1,...,αn)) be the probability density function of the continuous random variable Θ.

Let D denote a directional derivative in the direction ($u_1,\ldots,u_n$).

Let L = -log(f) with (α1,...,αn) evaluated at (X1,...,Xn) as the defining equation for the (X1,...Xn) vector (i.e. Minimize L or determine the critical point(s) of L with respect to (X1,...,Xn)).

Let S = Dlog(f).

Let COV = Cov(X1,...,Xn).

Let MSE_{ij} = E{(Xi-αi)(Xj-αj)} = Mean Squared Error.

The above definitions are utilized in seven important and interrelated results.

RESULT#1

$$COV = MSE + O(\sigma^4)$$

Proof:

$$COV_{ij} = E\{(Xi-E[Xi])(Xj-E[Xj])\}$$

$$= E\{(Xi-\alpha i+E[Xi-\alpha i])(Xj-\alpha j+E[Xj-\alpha j])\}$$

$$= E\{(Xi-\alpha i)(Xj-\alpha j)\} + E\{(Xi-\alpha i)E[Xj-\alpha j]\} + \ldots + E\{Xi-\alpha i\}E\{Xj-\alpha j\}$$

$$= MSE_{ij} + O(\sigma^4)$$

RESULT#2

The directional variance satisfies

$$(u_1,\ldots,u_n)\ MSE\ (u_1,\ldots,u_n)^T \quad \geq \quad \frac{1+O(\sigma^2)}{E\{(S)^2\}}$$

Equality holds if

$u\cdot(X1-\alpha 1, X2-\alpha 2,\ldots,Xn-\alpha n)$ is proportional to S

Proof:

$$\begin{aligned}(u_1,\ldots,u_n) &= D(\alpha 1,\ldots,\alpha n) = D\int\ldots\int (x1,\ldots,xn)\ f \ + D(O(\sigma^2),\ldots,O(\sigma^2))\\
&= D\int\ldots\int (x1,\ldots,xn)\ f - (\alpha 1,\ldots,\alpha n)D\int\ldots\int f\\
&\qquad + (O(\sigma^2),\ldots,O(\sigma^2))\\
&= \int\ldots\int (x1,\ldots,xn)\ D[f] - (\alpha 1,\ldots,\alpha n)\int\ldots\int D[f]\\
&\qquad + (O(\sigma^2),\ldots,O(\sigma^2))\\
&= \int\ldots\int [(x1,\ldots,xn)-(\alpha 1,\ldots,\alpha n)]\ D[f] + (O(\sigma^2),\ldots,O(\sigma^2))\\
&= E\{[(X1,\ldots,Xn)-(\alpha 1,\ldots,\alpha n)](S)\} + (O(\sigma^2),\ldots,O(\sigma^2))\end{aligned}$$

Taking the dot product on both sides with $(u_1,\ldots,u_n)$ and squaring,

$$\left|(u_1,\ldots,u_n)\cdot(u_1,\ldots,u_n) + O(\sigma^2)\right|^2$$
$$= [E\{[(u_1,\ldots,u_n)(X1-\alpha 1,\ldots,Xn-\alpha n)](S)\}]^2$$

Now, by the Cauchy-Schwarz inequality

$$\left|(u_1,\ldots,u_n)\cdot(u_1,\ldots,u_n) + O(\sigma^2)\right|^2$$
$$\leq E\{[(u_1,\ldots,u_n)(X1-\alpha 1,\ldots,Xn-\alpha n)]^2\}E\{(S)^2\}$$
$$= (u_1,\ldots,u_n)\ MSE\ (u_1,\ldots,u_n)^T\ \ E\{(S)^2\}$$

DEFINITIONS AND ASSUMPTIONS APPLYING TO THE NORMAL CASE:

Assume the Θ_i are independently Normally distributed with mean ,

$\theta_i(\alpha 1,\ldots,\alpha n)$ and standard deviation, σ^2.

Further, assume that the function associated with the mean above may be

computed at the estimate, i.e. $\theta_i(X1,\ldots,Xn)$.

Finally, assume that Xi evaluated at $(\alpha_1,\ldots,\alpha_n)$ is α_i (i.e. The true parameter is computed if there is no 'observation error' and the mean observation is the no error case).

Let $\varepsilon_i = \theta_i - \theta_i(\alpha 1,\ldots,\alpha n)$. In some cases, this may be interpreted as error.

Let $g_i(\varepsilon_1\ldots\varepsilon_n)=Xi(\theta_1,\theta_2\ldots\theta_n)=Xi(\varepsilon_1+\theta_1(\alpha 1\ldots\alpha n),\ldots,\varepsilon_n+\theta_n(\alpha 1\ldots\alpha n))$

g_{i,ε_k} denotes the partial derivative with respect to ε_k evaluated at $\varepsilon_i=0$ for all i.

$\theta_{i,\alpha j}$ denotes the partial derivative with respect to αj.

MSE_L represents the Mean Standard Error matrix for the Least Squares approach (i.e. Minimizing S^2).

TAYLOR SERIES EXPANSION

Expanding $\{g_1(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_m),\ldots,g_n(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_m)\}$

$$= (\alpha 1,\ldots,\alpha n)+\Sigma[g_{1,\varepsilon_i},\ldots,g_{n,\varepsilon_i}]\varepsilon_i +\Sigma\Sigma O(\varepsilon_i\varepsilon_j).$$

RESULT#3

Let θ_i have an independent Normal distribution with mean, $\theta_i(\alpha 1,\ldots,\alpha n)$ and variance, σ^2.

$$E\{(S)^2\} = [1/\sigma^2]\ \Sigma\ [u_1\ldots u_n] \begin{vmatrix} (\theta_{i,\alpha 1})^2 & \cdots & \theta_{i,\alpha 1}\theta_{i,\alpha n} \\ \vdots & \ddots & \vdots \\ \theta_{i,\alpha n}\theta_{i,\alpha 1} & \cdots & (\theta_{i,\alpha n})^2 \end{vmatrix} [u_1\ldots u_n]^T$$

Proof:

$$\begin{aligned} S &= \Sigma\{([\theta_i-\theta_i(\alpha 1\ldots\alpha n)]/\sigma^2)\theta_{i,\alpha 1}\ u_1 - \ldots \\ &\quad + \Sigma\{([\theta_i-\theta_i(\alpha 1\ldots\alpha n)]/\sigma^2)\theta_{i,\alpha n}\ u_n \\ &= \Sigma\{([\theta_i-\theta_i(\alpha 1\ldots\alpha n)]/\sigma^2)(\theta_{i,\alpha 1}\ u_1 + \ldots + \theta_{i,\alpha n}\ u_n)\} \end{aligned}$$

Therefore,

$E\{(S)^2\}$

$= E\{\Sigma\Sigma[\epsilon_i/\sigma^2][\epsilon_j/\sigma^2][\theta_{i,\alpha 1}\ u_1 + \ldots + \theta_{i,\alpha n}\ u_n][\theta_{j,\alpha 1}\ u_1 + \ldots + \theta_{j,\alpha n}\ u_n]\}$

$= [1/\sigma^4][\Sigma\Sigma E\{(\epsilon_i\epsilon_j)(\theta_{i,\alpha 1}\ u_1 + \ldots + \theta_{i,\alpha n}\ u_n)(\theta_{j,\alpha 1}\ u_1 + \ldots + \theta_{j,\alpha n}\ u_n)\}]$

$= [1/\sigma^2][\Sigma(\theta_{i,\alpha 1}\ u_1 + \ldots + \theta_{i,\alpha n}\ u_n)(\theta_{i,\alpha 1}\ u_1 + \ldots + \theta_{i,\alpha n}\ u_n)]$

RESULT#4

Let θ_i have an independent Normal distribution with mean, $\theta_i(\alpha 1,\ldots,\alpha n)$ and variance, σ^2.

$$MSE_L = [\sigma^2]\ \Sigma \begin{vmatrix} (\theta_{i,\alpha 1})^2 & \cdots & \theta_{i,\alpha 1}\theta_{i,\alpha n} \\ \vdots & \ddots & \vdots \\ \theta_{i,\alpha 1}\theta_{i,\alpha n} & \cdots & (\theta_{i,\alpha n})^2 \end{vmatrix}^{-1} + O(\sigma^4)$$

Proof:

$L = \Sigma(\theta_i - \theta_i(X1..Xn))^2/(2\sigma^2) + (n/2)\ln(2\pi\sigma^2)$

Recall that g_{k,ϵ_i} is evaluated at the true location.

The Mean Standard Error matrix is by definition

$$MSE = E\left\{\begin{vmatrix} (X1-\alpha 1)^2 & \cdots & (X1-\alpha 1)(Xn-\alpha n) \\ \vdots & \ddots & \vdots \\ (X1-\alpha 1)(Xn-\alpha n) & \cdots & (X1-\alpha n)^2 \end{vmatrix}\right\}$$

$$= E\{\Sigma\Sigma(g_{1,\epsilon_i},\ldots,g_{n,\epsilon_i})^T(g_{1,\epsilon_j},\ldots,g_{n,\epsilon_j})\epsilon_i\epsilon_j\} + O(\sigma^4)$$

$$= \Sigma(g_{1,\epsilon_i},\ldots,g_{n,\epsilon_i})^T(g_{1,\epsilon_i},\ldots,g_{n,\epsilon_i})\sigma^2 + O(\sigma^4)$$

Recall that $(g_1,\ldots,g_n)$ are the $(X1,\ldots,Xn)$ determined where $L_{X1} = \ldots = L_{Xn} = 0$.

Differentiating the defining equations with respect to ϵ_i

$$L_{X1}=0 \;\ldots\; L_{Xn}=0 \Longrightarrow \quad dL_{X1}/d\varepsilon_i = L_{X1X1}g_{1,\varepsilon_i}+\cdots+L_{X1Xn}g_{n,\varepsilon_i}+L_{X1\varepsilon_i}=0$$

$$\vdots$$

$$dL_{Xn}/d\varepsilon_i = L_{XnX1}g_{1,\varepsilon_i}+\cdots+L_{XnXn}g_{n,\varepsilon_i}+L_{Xn\varepsilon_i}=0$$

$$\begin{vmatrix} g_{1,\varepsilon_i} \\ \vdots \\ g_{n,\varepsilon_i} \end{vmatrix} = - \begin{vmatrix} L_{X1X1} & \cdots & L_{X1Xn} \\ \vdots & \ddots & \vdots \\ L_{X1Xn} & \cdots & L_{XnXn} \end{vmatrix}^{-1} \begin{vmatrix} L_{X1\varepsilon_i} \\ \vdots \\ L_{Xn\varepsilon_i} \end{vmatrix}$$

$$L_{Xj} = \sum_{t=1}^{n}(-2)(\theta_{t,Xj})(\theta_t(X1,\ldots,Xn)-\hat{\theta}_t)/(2\sigma^2)$$

$$L_{XjXk} = \sum_{t=1}^{n}(-2)[(\theta_{t,Xj})(\theta_{t,Xk})+(\theta_{t,XjXk})(\theta_t(X1,\ldots,Xn)-\hat{\theta}_t)]/(2\sigma^2)$$

$$= \sum_{t=1}^{n}(-2)(\theta_{t,Xj})(\theta_{t,Xk})/(2\sigma^2) \qquad (\text{Recall that } \varepsilon_i=0)$$

Since, $(2\sigma^2)L_{Xj} = \sum_{t=1}^{n}(-2)(\theta_{t,Xj})[\theta_t(X1,\ldots,Xn)-\theta_t(\alpha1,\ldots,\alpha n)+\theta_t(\alpha1,\ldots,\alpha n)-\hat{\theta}_t]$

$$= \sum_{t=1}^{n}(-2)(\theta_{t,Xj})[\theta_t(X1,\ldots,Xn)-\theta_t(\alpha1,\ldots,\alpha n)] + \sum_{t=1}^{n}2(\theta_{t,Xj})\varepsilon_t$$

$$L_{Xj\varepsilon_i} = -2\theta_{i,Xj}/(2\sigma^2)$$

Therefore,

$$\begin{vmatrix} g_{1,\varepsilon_i} \\ \vdots \\ g_{n,\varepsilon_i} \end{vmatrix} = \begin{vmatrix} 2\Sigma(\theta_{t,X1})^2 & \cdots & 2\Sigma(\theta_{t,X1})(\theta_{t,Xn}) \\ \vdots & \ddots & \vdots \\ 2\Sigma(\theta_{t,Xn})(\theta_{t,X1}) & \cdots & 2\Sigma(\theta_{t,Xn})^2 \end{vmatrix}^{-1} \begin{vmatrix} 2\theta_{i,X1} \\ \vdots \\ 2\theta_{i,Xn} \end{vmatrix}$$

Thus,

$$(g_{1,\varepsilon_i},\ldots,g_{n,\varepsilon_i})^T(g_{1,\varepsilon_i},\ldots,g_{1,\varepsilon_i})$$

$$= \begin{vmatrix} (\theta_{i,X1})^2 & \cdots & (\theta_{i,X1})(\theta_{i,Xn}) \\ \vdots & \ddots & \vdots \\ (\theta_{i,Xn})(\theta_{i,X1}) & \cdots & (\theta_{i,Xn})^2 \end{vmatrix}^{-1}$$

However, g_i is evaluated at the true, which when done produces the desired result

$$= \begin{vmatrix} (\theta_{i,\alpha 1})^2 & \cdots & (\theta_{i,\alpha 1})(\theta_{i,\alpha n}) \\ \vdots & \ddots & \vdots \\ (\theta_{i,\alpha n})(\theta_{i,\alpha 1}) & \cdots & (\theta_{i,\alpha n})^2 \end{vmatrix}^{-1}$$

RESULT#5

$$\bar{u}^{>} \text{ MSE } (\bar{u}^{>})^T \geq \frac{1 + O(\sigma^2)}{\bar{u}^{>} \text{ MSE}_L^{-1} (\bar{u}^{>})^T} \qquad \text{where } \bar{u}^{>} = (u_1, u_2, \ldots, u_n)$$

Proof:

From Result # 2

$$(u_1, \ldots, u_n) \text{ MSE } (u_1, \ldots, u_n)^T \quad \geq \quad \frac{1+O(\sigma^2)}{E\{(S)^2\}}$$

From Result # 3

$$E\{(S)^2\} = [1/\sigma^2]\Sigma\ [u_1 \ldots u_n] \begin{vmatrix} (\theta_{i,\alpha 1})^2 & \cdots & \theta_{i,\alpha 1}\theta_{i,\alpha n} \\ \vdots & \ddots & \vdots \\ \theta_{i,\alpha n}\theta_{i,\alpha 1} & \cdots & (\theta_{i,\alpha n})^2 \end{vmatrix} [u_1 \ldots u_n]^T$$

From Result # 4

$$\text{MSE}_L = [\sigma^2]\ \Sigma \begin{vmatrix} (\theta_{i,\alpha 1})^2 & \cdots & \theta_{i,\alpha 1}\theta_{i,\alpha n} \\ \vdots & \ddots & \vdots \\ \theta_{i,\alpha 1}\theta_{i,\alpha n} & \cdots & (\theta_{i,\alpha n})^2 \end{vmatrix}^{-1} + O(\sigma^4)$$

RESULT#6

Maximum eigenvalue of $\text{MSE}_L \leq$ Maximum eigenvalue for other methods

Proof:

Let λ_L be the maximum eigenvalue of the L method.

Let λ_{other} be the maximum eigenvalue of some method.

Let $\bar{v}^{>}$ be the eigenvector associated with λ_L.

$$\frac{1}{\bar{v}^{>} \quad MSE_L^{-1} \quad (\bar{v}^{>})^T} = \frac{1}{1/\lambda_L} = \lambda_L$$

From Result # 5, it follows that

$$\lambda_L \le \bar{v}^{>} MSE_{other}(\bar{v}^{>})^T$$

Finally, since MSE_{other} is a positive definite matrix, $\bar{v}^{>}MSE_{other}(\bar{v}^{>})^T$ is between the minimum and minimum eigenvalues for MSE_{other}. Thus,

$$\lambda_L \le \lambda_{other}$$

RESULT#7

Minimum eigenvalue of $MSE_L \le$ Minimum eigenvalue for other methods

Proof:

Let λ_{other} be the minimum eigenvalue of some method.

Let λ_L be the minimum eigenvalue of the L method.

Let $\bar{v}^{>}$ be the eigenvector associated with λ_{other}.

$$\lambda_L = \frac{1}{1/\lambda_L}$$

$$= \frac{1}{\max \quad \bar{v}^{>} \quad MSE_L^{-1} \quad (\bar{v}^{>})^T}$$

$$\le \bar{v}^{>} MSE_{other} (\bar{v}^{>})^T = \lambda_{other} \qquad \text{Recall Result \# 5}$$

Therefore,

$$\lambda_L \le \lambda_{other}$$

Observations - Comparing error ellipses between the L and any other estimator at the same confidence level we note:

1) The largest axis of the L error ellipse is smaller than the largest axis of the other estimator.
2) The smallest axis of the L error ellipse is smaller than the smallest axis of the other estimator.
3) If one is only in two dimensions, 1) and 2) are enough to imply that the L error ellipse is smaller in size.

In n dimensions, intermediate axes must also be considered in any discussion of error ellipse size, thus 1) and 2) are not sufficient to show that the L error ellipse is smaller in the general case.

END
DATE
FILMED
DTIC
9-88